

SEQUENZPROTEIN 534 Rec'd PCT/PT 03 JUL 2000

(1) ALGEMEINE INFORMATION:

(i) ANMELDER:

- (A) NAME: BASF Aktiengesellschaft
- (B) STRASSE: Carl Bosch Strasse
- (C) ORT: Ludwigshafen
- (D) BUNDESLAND: Rheinland-Pfalz
- (E) LAND: Germany
- (F) POSTLEITZAHL: D-67056

(ii) ANMELDETITEL: Orotidin-5'-Phosphatdecarboxylase-Gen,
Genkonstrukt enthaltend dieses Gen und seine Verwendung

(iii) ANZAHL DER SEQUENZEN: 2

(iv) COMPUTER-LESBARE FORM:

- (A) DATENTRÄGER: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) BETRIEBSSYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPA)

(2) INFORMATION ZU SEQ ID NO: 1:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 1380 Basenpaare
- (B) ART: Nukleinsäure
- (C) STRANGFORM: Einzel
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: DNS (genomisch)

(iii) HYPOTHETISCH: NEIN

(iii) ANTISENSE: NEIN

(vi) URSPRÜNGLICHE HERKUNFT:

- (A) ORGANISMUS: Ashbya gossypii

(vii) UNMITTELBARE HERKUNFT:

- (B) CLON: ura3

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: CDS
- (B) LAGE: 210..1013

(ix) MERKMALE:

- (A) NAME/SCHLÜSSEL: 5'UTR
- (B) LAGE: 1..199

(ix) MERKMALE:

(A) NAME/SCHLÜSSEL: 3'UTR

(B) LAGE: 1014..1380

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

CTCGAGCAAC TCATTGGAAG CCCTTCGCAA ACGACCTCTA TATCTCGTCT CAAGTTCCTA 60

CTATCATGTA TGCTGTCACT ACAGAAAAAT TTTTGTCTAT AGCTGGCAAG AAGCACATCA 120

CATACATTCT GATGGTGTAG GCTCCACATC ACAGTAAGCA TTTGTATAAG GCTGATCACA 180

TAGGGTGCTA CCGACCTAGC CATTGCCAC ATG TCA ACG AAA TCT TAC GCA GAA 233

Met Ser Thr Lys Ser Tyr Ala Glu

1 5

AGG GCC AAG GCA CAC AAT TCG CCA GTT GCT AGA AAG CTT CTG GCA TTG 281

Arg Ala Lys Ala His Asn Ser Pro Val Ala Arg Lys Leu Leu Ala Leu

10 15 20

ATG CAC GAG AAG AAA ACC AAT CTC TGC GCT TCC CTT GAT GTG CGG ACG 329

Met His Glu Lys Lys Thr Asn Leu Cys Ala Ser Leu Asp Val Arg Thr

25 30 35 40

TCT AGA AAG CTT CTG GAG CTA GCA GAC ACG CTG GGA CCG CAC ATT TGT 377

Ser Arg Lys Leu Leu Glu Leu Ala Asp Thr Leu Gly Pro His Ile Cys

45 50 55

CTG CTG AAG ACA CAT GTC GAC ATA CTG ACG GAC TTC GAC ATC GAG ACG 425

Leu Leu Lys Thr His Val Asp Ile Leu Thr Asp Phe Asp Ile Glu Thr

60 65 70

ACA GTC AAG CCG CTG CAG CAG CTT GCG GCT AAG CAC AAC TTC ATG ATC 473

Thr Val Lys Pro Leu Gln Gln Leu Ala Ala Lys His Asn Phe Met Ile

75 80 85

TTC GAG GAC CGC AAG TTC GCT GAC ATT GGC AAC ACG GTT AAG CTG CAG 521

Phe Glu Asp Arg Lys Phe Ala Asp Ile Gly Asn Thr Val Lys Leu Gln

90 95 100

TAC TCC TCC GGC GTG TAC CGT ATC GCG GAG TGG GCG GAT ATT ACC AAT 569

Tyr Ser Ser Gly Val Tyr Arg Ile Ala Glu Trp Ala Asp Ile Thr Asn

105 110 115 120

GCA CAC GGC GTC ACC GGC CCC GGT GTG ATA GCC GGG CTG AAG GAG GCT 617

Ala His Gly Val Thr Gly Pro Gly Val Ile Ala Gly Leu Lys Glu Ala

125 130 135

GCG AAA CTG GCC TCA CAG GAA CCC AGG GGG TTG CTG ATG CTG GCA GAG 665

Ala Lys Leu Ala Ser Gln Glu Pro Arg Gly Leu Leu Met Leu Ala Glu

140 145 150

0058279 " 070300

CTC	TCT	TCT	CAG	GGC	TCT	TTG	GCG	CGC	GGA	GAC	TAT	ACC	GCG	GGC	GTC	713
Leu	Ser	Ser	Gln	Gly	Ser	Leu	Ala	Arg	Gly	Asp	Tyr	Thr	Ala	Gly	Val	
		155					160					165				
GTT	GAA	ATG	GCG	AAG	CTG	GAC	GAA	GAC	TTT	GTG	ATC	GGG	TTC	ATC	GCG	761
Val	Glu	Met	Ala	Lys	Leu	Asp	Glu	Asp	Phe	Val	Ile	Gly	Phe	Ile	Ala	
	170					175					180					
CAG	CGT	GAC	ATG	GGT	GGG	CGT	GCA	GAC	GGC	TTT	GAC	TGG	CTC	ATC	ATG	809
Gln	Arg	Asp	Met	Gly	Gly	Arg	Ala	Asp	Gly	Phe	Asp	Trp	Leu	Ile	Met	
185					190				195						200	
ACC	CCG	GGG	GTT	GGC	CTG	GAC	GAC	AAA	GGA	GAC	GGC	CTG	GGC	CAG	CAG	857
Thr	Pro	Gly	Val	Gly	Leu	Asp	Asp	Lys	Gly	Asp	Gly	Leu	Gly	Gln	Gln	
			205					210					215			
TAC	CGC	ACG	GTG	GAT	GAG	GTC	GTC	AGC	GAC	GGT	ACC	GAT	GTG	ATC	ATT	905
Tyr	Arg	Thr	Val	Asp	Glu	Val	Val	Ser	Asp	Gly	Thr	Asp	Val	Ile	Ile	
		220					225					230				
GTT	GGC	AGA	GGG	CTC	TTT	GAC	AAG	GGA	AGA	GAC	CCC	AAG	GTC	GAG	GGT	953
Val	Gly	Arg	Gly	Leu	Phe	Asp	Lys	Gly	Arg	Asp	Pro	Lys	Val	Glu	Gly	
	235					240				245						
GCC	CGC	TAC	CGC	AAG	GCC	GGT	TGG	GAG	GCT	TAC	TTG	CGC	CGT	ATG	GGC	1001
Ala	Arg	Tyr	Arg	Lys	Ala	Gly	Trp	Glu	Ala	Tyr	Leu	Arg	Arg	Met	Gly	
	250					255				260						
GAG	ACT	TCG	TAGTCTATCG	CTGGCGCCCA	CAGTATATAG	GCGGATTCCA										1050
Glu	Thr	Ser														
265																
CCGCCGATTA	CCATCTCAGC	AACCTTTTGTG	TAATTATATG	CCCCTATTGC	CCTTATTTCC											1110
GAGCTGGTGC	CGGGATCGGT	TTATAGACGG	GCAACAAGTT	GATACTTTGT	TCAGTAGCAT											1170
GCATCCAACA	CTTGCAGGCT	TGGGGTGTGG	AAGGCCTCGC	CGCGGATAAT	TCGTATTACC											1230
CGCACTTCGT	GAAGTATTGC	TTTATGAAAA	ATCTTCACTT	TGGGCTAACT	AGAGCCATAA											1290
CTCGACACAA	GCCCCTTCCT	ACACACTTCG	AGCTGGGACT	AAAGTGACAA	CGAATAGCAA											1350
ATAATTAGCA	AATATGGATG	CGTTGAATTC														1380

(2) INFORMATION ZU SEQ ID NO: 2:

(i) SEQUENZ CHARAKTERISTIKA:

- (A) LÄNGE: 267 Aminosäuren
- (B) ART: Aminosäure
- (D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Protein

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

Met Ser Thr Lys Ser Tyr Ala Glu Arg Ala Lys Ala His Asn Ser Pro
 1 5 10 15
 Val Ala Arg Lys Leu Leu Ala Leu Met His Glu Lys Lys Thr Asn Leu
 20 25 30
 Cys Ala Ser Leu Asp Val Arg Thr Ser Arg Lys Leu Leu Glu Leu Ala
 35 40 45
 Asp Thr Leu Gly Pro His Ile Cys Leu Leu Lys Thr His Val Asp Ile
 50 55 60
 Leu Thr Asp Phe Asp Ile Glu Thr Thr Val Lys Pro Leu Gln Gln Leu
 65 70 75 80
 Ala Ala Lys His Asn Phe Met Ile Phe Glu Asp Arg Lys Phe Ala Asp
 85 90 95
 Ile Gly Asn Thr Val Lys Leu Gln Tyr Ser Ser Gly Val Tyr Arg Ile
 100 105 110
 Ala Glu Trp Ala Asp Ile Thr Asn Ala His Gly Val Thr Gly Pro Gly
 115 120 125
 Val Ile Ala Gly Leu Lys Glu Ala Ala Lys Leu Ala Ser Gln Glu Pro
 130 135 140
 Arg Gly Leu Leu Met Leu Ala Glu Leu Ser Ser Gln Gly Ser Leu Ala
 145 150 155 160
 Arg Gly Asp Tyr Thr Ala Gly Val Val Glu Met Ala Lys Leu Asp Glu
 165 170 175
 Asp Phe Val Ile Gly Phe Ile Ala Gln Arg Asp Met Gly Gly Arg Ala
 180 185 190
 Asp Gly Phe Asp Trp Leu Ile Met Thr Pro Gly Val Gly Leu Asp Asp
 195 200 205
 Lys Gly Asp Gly Leu Gly Gln Gln Tyr Arg Thr Val Asp Glu Val Val
 210 215 220
 Ser Asp Gly Thr Asp Val Ile Ile Val Gly Arg Gly Leu Phe Asp Lys
 225 230 235 240
 Gly Arg Asp Pro Lys Val Glu Gly Ala Arg Tyr Arg Lys Ala Gly Trp
 245 250 255
 Glu Ala Tyr Leu Arg Arg Met Gly Glu Thr Ser
 260 265